It is claimed:

1. A computer-implemented method for building an artificial neural network, wherein the artificial neural network predicts at least one target based upon observations defined in a state space, comprising the steps of:

retrieving an input data set that includes the observations and the target;

inserting in the state space a plurality of points based upon the values of the observations in the state space, wherein the number of inserted points is less than the number of observations;

determining a statistical measure that describes a relationship between the observations and the inserted points; and

determining weights and activation functions of the artificial neural network using the statistical measure.

2. The method of claim 1 further comprising the steps of:

determining within the state space a range based upon the observations; and

inserting within the range the points for use in determining the weights and activation functions of the artificial neural network.

3. The method of claim 2 wherein the inserted points have initial and final points within the range, said method further comprising the step of:

determining the points within the range such that the initial and final points are spaced farther apart from their neighboring points than points located in the middle of the range.

4. The method of claim 1 further comprising the steps of:

determining principal components for the input data set;

selecting the principal components that are substantially associated to the target.

5. The method of claim 4 further comprising the step of:

determining which of the principal components are substantially correlated by using a linear regression model.

6. The method of claim 5 further comprising the step of:

determining which of the principal components are substantially correlated by using R-Square values from the linear regression model.

7. The method of claim 5 further comprising the step of:

determining which of the principal components are substantially correlated by using F values from the linear regression model.

8. The method of claim 2 wherein the statistical measure is a frequency statistical measure, said method further comprising the step of:

generating a frequency table that describes a relationship between score values of the selected principal components and the inserted points.

9. The method of claim 8 wherein the frequency table has a dimensional size that is based upon the number of selected principal components and the number of inserted points.

10. The method of claim 8 wherein the frequency table contains cells that are defined by ranges between the inserted points, wherein each of the cells contains a counter that counts number of observations for which a selected principal component has a value within a cell's range.

11. The method of claim 10 wherein the sum of entries on a row of the frequency table equals the number of observations in the input data set.

12. The method of claim 1 wherein candidate activation functions are available for use as the activation functions of the artificial neural network, wherein the candidate activation functions differ in function type from each other.

13. The method of claim 12 wherein a first candidate activation function type is selected within a first layer of the artificial neural network, wherein a second candidate activation function type is selected within a second layer of the artificial neural network, wherein the first candidate activation function type is a different function type than the second candidate activation function type.

14. The method of claim 12 further comprising the steps of:

determining principal components for the input data set;

selecting the principal components that are substantially correlated to the target; and

generating a frequency table that describes frequency relationships between score values of the selected principal components and the inserted points.

15. The method of claim 14 further comprising the step of:

determining which of the candidate activation functions to use within a layer of the artificial neural network by using the frequency relationships.

16. The method of claim 14 further comprising the steps of:

determining parameters of the candidate activation functions by optimizing the candidate activation functions with respect to a predetermined objective function;

selecting which of the candidate activation functions to use within a layer of the artificial neural network; and

creating a layer of the artificial neural network with the selected candidate activation function and its respective optimized parameters.

17. The method of claim 16 wherein the objective function is a sum of squares error objective function.

18. The method of claim 16 wherein the objective function is an accuracy rate objective function.

19. The method of claim 16 wherein a layer weight is determined during the optimizing of the candidate activation functions.

20. The method of claim 16 wherein the frequency table specifies which observations of the selected principal components is assigned a greater weight during the optimization of the candidate activation functions.

21. The method of claim 16 further comprising the steps of:

generating prediction outcomes for each of the candidate activation functions; and

selecting one of the candidate activation functions to use within a layer of the artificial neural network based upon the generated prediction outcomes.

22. The method of claim 21 wherein the optimized parameters of the candidate activation functions are used to generate the prediction outcomes.

23. The method of claim 21 wherein the prediction outcomes are generated by testing each of the candidate activation functions with the principal components and the observations.

24. The method of claim 21 wherein the observations are passed through a linking web into each of the candidate activation functions to evaluate fit of the prediction outcomes to an evaluation data set.

25. The method of claim 24 wherein the input data set is used as the evaluation data set for determining a first stage of the artificial neural network.

26. The method of claim 25 wherein residuals from the first stage are used as the evaluation data set for determining a second stage of the artificial neural network.

27. The method of claim 26 wherein residuals from the second stage are used as the evaluation data set for determining a third stage of the artificial neural network.

28. The method of claim 21 wherein the parameters of the candidate activation functions are generated substantially in parallel.

29. The method of claim 21 wherein the prediction outcomes for the candidate activation functions are generated substantially in parallel.

30. The method of claim 1 wherein stages of the artificial neural network are determined until a predetermined number of stages are achieved.

31. The method of claim 1 wherein stages of the artificial neural network are determined until the artificial neural network's predictive capability does not improve within a predetermined amount.

32. The method of claim 1 wherein stages of the artificial neural network are determined until the artificial neural network's predictive capability reaches a level of prediction that satisfies a predetermined threshold.

33. The method of claim 1 wherein the input data set is pre-processed before the points are inserted in the state space.

34. The method of claim 33 wherein the input data set includes class variables, wherein the input data set is pre-processed by generating dummy variables from class variables.

35. The method of claim 33 wherein the input data set is pre-processed by being normalized.

36. The method of claim 33 wherein the input data set includes an interval target, wherein the input data set is pre-processed by having the interval target be decile ranked.

37. The method of claim 1 wherein dimensions of the state space are defined by the observations and the target.

38. The method of claim 1 wherein the built artificial neural network predicts multiple targets based upon the observations.

39. A computer-implemented method for building an artificial neural network from a set of candidate activation functions, comprising the steps of:

retrieving an input data set that includes observations and at least one target for the observations;

reducing the input data set such that the reduced input data set contains a number of points less than the number of observations;

optimizing parameters of the candidate activation functions with respect to the reduced input data set through use of an objective function;

generating results for each of the candidate activation functions using the optimized parameters of the candidate activation functions and the reduced input data set;

selecting a first activation function from the candidate activation functions based upon the generated simulated results;

using the selected first activation function within a first layer of the artificial neural network,

wherein residuals result from predictions by the first layer's selected activation function of the target; and

selecting a second activation function from the candidate activation functions to form a second layer based upon the second activation function's capability to predict the residuals.

40. The method of claim 39 wherein the candidate activation functions differ in function type from each other.

41. The method of claim 40 wherein a first candidate activation function type is used within a first layer of the artificial neural network, wherein a second candidate activation function type is used within a second layer of the artificial neural network, wherein the first candidate activation function type is a different function type than the second candidate activation function type.

42. The method of claim 40 further comprising the steps of:

determining principal components for the input data set;

selecting the principal components that are substantially correlated to the target; and

generating a frequency table that describes frequency relationships between values of the selected principal components and the inserted points.

43. The method of claim 42 further comprising the step of:

determining which of the candidate activation functions to use within a layer of the artificial neural network by using the frequency relationships.

44. The method of claim 42 further comprising the steps of:

determining parameters of the candidate activation functions by optimizing the candidate activation functions with respect to a predetermined objective function;

selecting which of the candidate activation functions to use within a layer of the artificial neural network; and

creating a layer of the artificial neural network with the selected candidate activation function and its respective optimized parameters.

45. The method of claim 44 wherein the objective function is a sum of squares error objective function.

46. The method of claim 44 wherein the objective function is an accuracy rate objective function.

47. The method of claim 44 wherein a layer weight is determined during the optimizing of the candidate activation functions.

48. The method of claim 44 wherein the frequency table specifies which observations of the selected principal components is accorded a greater weight during the optimizing of the candidate activation functions.

49. The method of claim 44 further comprising the steps of:

generating prediction outcomes for each of the candidate activation functions; and

selecting one of the candidate activation functions to use within a layer of the artificial neural network based upon the generated prediction outcomes.

50. The method of claim 49 wherein the optimized parameters of the candidate activation functions are used to generate the prediction outcomes.

51. The method of claim 49 wherein the prediction outcomes are generated by testing each of the candidate activation functions with the principal components and the observations.

52. The method of claim 49 wherein the observations are passed through a linking web into each of the candidate activation functions to evaluate fit of the prediction outcomes to an evaluation data set.

53. The method of claim 52 wherein the input data set is used as the evaluation data set for determining a first stage of the artificial neural network.

54. The method of claim 53 wherein residuals from the first stage are used as the evaluation data set for determining a second stage of the artificial neural network.

55. The method of claim 54 wherein residuals from the second stage are used as the evaluation data set for determining a third stage of the artificial neural network.

56. The method of claim 49 wherein the parameters of the candidate activation functions are generated substantially in parallel.

57. The method of claim 49 wherein the prediction outcomes for the candidate activation functions are generated substantially in parallel.

58. An artificial neural network that predicts at least one target based upon observations defined in a state space, comprising:

a first stage that contains a first activation function type, wherein the first layer is predictive of the target, wherein residuals result from predictions by the first stage of the target; and

a second stage that contains a second activation function type, wherein the second layer is predictive of the residuals resulting from the predictions by the first stage.

59. The artificial neural network of claim 58 wherein the first activation function type is a nonlinear function.

60. The artificial neural network of claim 59 wherein the first activation function type includes at least one parameter that is optimized using a nonlinear optimization technique.

61. The artificial neural network of claim 60 wherein the nonlinear optimization technique is the Levenberg-Marquardt optimization.

62. The artificial neural network of claim 58 further comprising an orthogonal layer that calculates an uncorrelated input data set from the observations.

63. The artificial neural network of claim 62 wherein the orthogonal layer determines principal components using principal component analysis.

64. The artificial neural network of claim 63 wherein the orthogonal layer computes a subset of the principal components to determine the input to the first layer.

65. The artificial neural network of claim 64 wherein the subset of principal components is determined by maximizing a measure of association among a set of principal components and the target.

66. The artificial neural network of claim 65 wherein the subset of principal components includes the principal components for which a sum of square errors of the linear regression is minimum.

67. The artificial neural network of claim 65 wherein the subset of principal components includes the principal components for which an accuracy rate is maximum.

68. The artificial neural network of claim 58 wherein the first activation function type is determined from a statistical measure of the observations and the target.

69. The artificial neural network of claim 68 wherein the statistical measure is a frequency table populated by the observations.